Journal of Educational Psychology

A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games

Pieter Wouters, Christof van Nimwegen, Herre van Oostendorp, and Erik D. van der Spek

Online First Publication, February 4, 2013. doi: 10.1037/a0031311

Journal of Educational Psychology

0022-0663/13/$12.00 DOI: 10.1037/a0031311

A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games

Pieter Wouters, Christof van Nimwegen, and Herre van Oostendorp
Utrecht University

Erik D. van der Spek
Eindhoven University of Technology

It is assumed that serious games influences learning in 2 ways, by changing cognitive processes and by affecting motivation. However, until now research has shown little evidence for these assumptions. We used meta-analytic techniques to investigate whether serious games are more effective in terms of learning and more motivating than conventional instruction methods (learning: $k = 77$, $N = 5{,}547$; motivation: $k = 31$, $N = 2{,}216$). Consistent with our hypotheses, serious games were found to be more effective in terms of learning ($d = 0.29$, $p < .01$) and retention ($d = 0.36$, $p < .01$), but they were not more motivating ($d = 0.26$, $p > .05$) than conventional instruction methods. Additional moderator analyses on the learning effects revealed that learners in serious games learned more, relative to those taught with conventional instruction methods, when the game was supplemented with other instruction methods, when multiple training sessions were involved, and when players worked in groups.

Keywords: serious games, game-based learning, cognition, motivation, meta-analysis

In the last decade, researchers have propagated the use of computer games for the purpose of learning and instruction (often referred to as serious games or game-based learning). In this respect, serious games are hypothesized to address both the cognitive and the affective dimensions of learning (O'Neil, Wainess, & Baker, 2005), to enable learners to adapt learning to their cognitive needs and interests, and to provide motivation for learning (Malone, 1981). Reviews regarding the effects of serious games show ambiguous results (Ke, 2009; Sitzmann, 2011; Vogel et al., 2006; Wouters, van der Spek, & van Oostendorp, 2009), but several scholars have noted that in general the quality of game research is poor (O'Neil et al., 2005) and that serious games are not more effective in terms of learning than other instruction methods when they are tested scientifically (Clark, Yates, Early, & Moulton, 2010). Many claims are supported by anecdotal arguments and lack sound empirical evidence. However, in the last 5 years, more well-designed empirical studies investigating the effects of serious games on learning and motivation have been published.

Our goal in this study was to statistically summarize the research on the effects of serious games on learning and motivation. Mayer (2011) has divided game research into three categories: a value-added approach, which questions how specific game features foster learning and motivation; a cognitive consequences approach, which investigates what people learn from serious games; and a media comparison approach, which investigates whether people learn better from serious games than from conventional media. Our meta-analysis adopted the media comparison approach. We compared serious games with conventional instruction methods such as lectures, reading, drill and practice, or hypertext learning environments. In addition, this study discerned instructional and contextual factors that may moderate the effectiveness and motivational appeal of serious games. Several meta-analyses have been conducted with respect to the effects of serious games (Ke, 2009; Sitzmann, 2011;Vogel et al., 2006). The meta-analysis by Ke (2009) is an interesting exploration of the field of game-based learning, but it does not statistically summarize effect sizes. The Vogel et al. (2006) meta-analysis investigated both cognitive and attitudinal effects and found that computer games and interactive simulations yielded higher cognitive outcomes than did conventional learning methods. Our meta-analysis expanded this research by incorporating the high number of well-designed studies that have been published in recent years and by focusing on other instructional and contextual factors, such as the number of training sessions with serious games and the moment of measurement of the learning effects (immediate or delayed). The more recent meta-analysis by Sitzmann (2011) focuses on simulation games, whereas our research has a broader perspective on serious games. Although this study shares some moderator variables with

Pieter Wouters, Christof van Nimwegen, and Herre van Oostendorp, Institute of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands; Erik D. van der Spek, Department of Industrial Design, Eindhoven University of Technology, Eindhoven, the Netherlands.

This research was funded by the Netherlands Organization for Scientific Research (Project No. 411-10-902). This research was also supported by the GATE project, funded by the Netherlands Organization for Scientific Research and the Netherlands ICT Research and Innovation Authority (ICT Regie).

Correspondence concerning this article should be addressed to Pieter Wouters, Utrecht University, Institute of Information and Computing Sciences, P.O. Box 80.089, 3508 TB Utrecht, the Netherlands. E-mail: P.J.M.Wouters@uu.nl

the Sitzmann study, we introduce new variables such as the domain in which the serious game is used, the age of the learners, and the group size (individual vs. group).

In the following sections we first define serious games. Next, we describe the theoretical framework with the main hypotheses and the moderator variables. The Method section comprises a description of the literature research, the inclusion criteria, the coding of the moderator variables, and the calculation of effect sizes. The Results section presents the general characteristics of the analysis, the main effects, and the effects of the moderator variables. Finally, we discuss the findings, draw conclusions, and depict some avenues for future research.

Definition of Serious Games

Several scholars have provided definitions or classifications of computer games characteristics (Garris, Ahlers, & Driskell, 2002; Malone, 1981; Prensky, 2001). For the purpose of this meta-analysis, we describe computer games in terms of being *interactive* (Prensky, 2001; Vogel et al., 2006), based on a set of *agreed rules and constraints* (Garris et al., 2002), and directed toward a clear *goal* that is often set by a *challenge* (Malone, 1981). In addition, games constantly provide *feedback,* either as a score or as changes in the game world, to enable players to monitor their progress toward the goal (Prensky, 2001). Some scholars contend that computer games also involve a competitive activity (against the computer, another player, or oneself), but it can be questioned if this is essentially a defining characteristic. Of course, there are many games in which the player is in competition with another player or with the computer, but in a game such as SimCity, players may actually enjoy the creation of a prosperous city that satisfies their beliefs or ideas without having the notion that they engage in a competitive activity. In the same vein, a narrative or the development of a story can be very important in a computer game (e.g., in adventure games), but again it is not a prerequisite for being a computer game (e.g., action games do not really require a narrative). In speaking of a serious (computer) game, we mean that the objective of the computer game is not to entertain the player, which would be an added value, but to use the entertaining quality for training, education, health, public policy, and strategic communication objectives (Zyda, 2005).

Theoretical Framework

In theory, games may influence learning in two ways, by changing the cognitive processes and by affecting the motivation. The (inter)active nature of computer game aligns with the current emphasis in educational psychology that active cognitive processing of educational material is a prerequisite for effective and sustainable learning (cf. Wouters, Paas, & van Merriënboer, 2008). Second, it is possible with computer games to simulate tasks in such a way that performing them in the game involves the same cognitive processes that are required for task performance in the real world (Tobias, Fletcher, Dai, & Wind, 2011). Finally, the immediate feedback in computer games provides players information regarding the correctness of their actions and decisions and thus gives them the opportunity to correct inaccurate information (Cameron & Dwyer, 2005; Moreno & Mayer, 2005).

Several classifications have been proposed for learning outcomes (for an overview, see Kraiger, Ford, & Salas, 1993; Wouters et al., 2009). In this meta-analysis, we focus on the cognitive dimension of learning. In the Wouters et al. (2009) classification, this dimension is divided into knowledge and cognitive skills. Knowledge refers to encoded knowledge reflecting both text-oriented learning (e.g., verbal knowledge) and non-text-oriented learning (e.g., knowledge derived from an image). A cognitive skill pertains to more complex cognitive processes, such as in problem solving when a learner applies knowledge and rules to achieve a solution for a (novel) situation. With reference to the aforementioned arguments, we will make a distinction in learning between knowledge and cognitive skills.

Our first hypothesis contends

1. That instruction with serious games yields higher learning gains than conventional instruction methods.

In the majority of studies, learning is measured immediately after the learning stage. The question can be raised whether such an immediate test is appropriate when the focus is on sustainable learning, which occurs when learners are still able to adequately apply the learned knowledge and skills in the long term. It is still an exception to include a delayed test in experimental designs. In this meta-analysis, sufficient pairwise comparisons were available to justify the inclusion of retention as a learning variable. For simulation games, there is some support that the acquired knowledge and skills are maintained over time (Pierfy, 1977; Sitzmann, 2011; van der Spek, 2011). In line with these results, we expect

2. That instruction with serious games will yield a higher level of retention than conventional instructional methods.

Several theories emphasize the potential of serious games to positively influence intrinsic motivation (Garris et al., 2002; Malone, 1981). This means that players are willing to invest more time and energy in game play not because of extrinsic rewards but because the game play in itself is rewarding. Several characteristics of serious games have been identified for this motivating appeal. Malone (1981) proposed that the most important factors that make playing a computer game intrinsically motivating are challenge, curiosity, and fantasy. Two other essential factors associated with computer games, autonomy (i.e., the opportunity to make choices) and competence (i.e., a task is experienced as challenging but not too difficult), originate from self-determination theory and are known to positively influence the experienced motivation (Przybylski, Rigby, & Ryan, 2010; Ryan, Rigby, & Przybylski, 2006). We therefore hypothesize

3. That instruction with serious games is more motivating than conventional instruction methods.

Moderator Variables for Learning

We also investigate how situational and contextual variables may moderate learning with serious games. We distinguish between hypothesized moderators, nonhypothesized moderators, and methodological moderators. First, we describe and ground the hypothesized moderators.

Learning arrangement of the comparison group. Modern educational theories advocate active cognitive processing as a prerequisite for genuine learning (Chi, de Leeuw, Chiu, & LaVancher, 1994; Mayer, 2001; Wittrock, 1974). In observational learning, for example, stronger learning effects are reported when learners engage in active coding (Bandura, 1976). Also, the research literature on self-explanations indicates that an active engagement of learners in the learning process fosters a better integration of new knowledge with prior knowledge and higher levels of transfer (Chi et al., 1994; Renkl & Atkinson, 2002; Roy & Chi, 2005). In this respect, a learning environment that stimulates an active cognitive attitude of the learners (e.g., doing practices and exercises) may foster more effective learning than does an arrangement in which learners are not explicitly prompted to actively engage in learning (e.g., reading an expository text or following a lecture). Therefore, the treatment that the comparison group receives may be an important moderator. We hypothesize that

4. The beneficial effect of serious games on learning is larger when the comparison group receives passive instruction than when the comparison group receives active instruction.

Serious game combined with other instructional methods. In computer games, players typically act and see the outcome of their actions reflected in changes in the game world. This may lead to a kind of intuitive learning: They know how to apply knowledge, but they cannot explicate it (Leemkuil & de Jong, 2011). Yet, it is important that learners verbalize their knowledge, because it enables them to integrate new knowledge with their prior knowledge, resulting in better recall and higher transfer of learning (Wouters et al., 2008). It is possible that supplemental instruction methods (e.g., discussion, explicit practice) enable learners to engage in learning activities that further support the articulation of knowledge.

Some evidence comes from Sitzmann (2011), who found that arrangements in which a simulation game was supplemented with other instructional methods yielded higher levels of learning. In arrangements in which only a simulation game was used, the comparison group performed better. In line with this observation, we hypothesize that

5. Relative to the comparison group, learning arrangements in which serious games are supplemented with other instruction methods will yield higher learning gains than will arrangements in which serious games are the only instruction method.

Number of training sessions. The question can be raised whether a training of only one session is sufficient to ensure cognitive changes. Serious games can be complex learning environments for several reasons. For example, players may have to attend to different locations on the screen and coordinate this with mouse or joystick movements, or they may have to engage in a task in which multiple variables that mutually interact play a role. It is plausible that, in comparison to that of conventional instruction methods, the effectiveness of serious games in terms of learning pays off only after multiple training sessions in which the players get used to the game. We hypothesize that

6. Multiple training sessions with serious games will yield higher learning gains than will multiple training sessions with conventional instruction methods.

Group size. One argument for collaborative learning in computer games is that it supports learners in articulating the knowledge that would otherwise have remained intuitive (van der Meij, Albers, & Leemkuil, 2011), but research comparing collaborative and solitary game play is ambiguous. The observation by Inkpen, Booth, Klawe, and Upitis (1995) that collaborative play resulted in significantly higher scores on motivation and learning outcomes than did solitary play was not confirmed by van der Meij et al. (2011). The meta-analysis by Vogel et al. (2006) revealed that both single users and groups showed higher cognitive gains in interactive simulations and games than in conventional teaching methods, but the effect size for single users was much larger than for groups. On the basis of these observations, we hypothesize that

7. Compared with the comparison group, single users will yield higher learning gains than will players who play in a group.

In addition to investigating these hypothesis-oriented moderator variables, we investigated other variables that potentially may have a moderating effect on learning. These include instructional domain, age of the player, the level of realism, and the use of a narrative.

Instructional domain. Serious games are used in different domains, ranging from domains that are part of school curricula (e.g., biology, mathematics), job-oriented domains (e.g., military), to more basic cognitive processing (e.g., visual attention). Some domains may be more connected with learning with serious games than are other domains.

Age. The question can be raised whether age is a moderator. The meta-analysis by Vogel et al. (2006), however, did not find differences between age groups in learning with serious games. In the light of the large number of studies over the last 5 years, we addressed this question again.

Level of realism. Designers of serious games have neither the money nor the time to create computer games that can match commercial computer games in level of realism. It is sometimes argued that players have expectations about the design of serious games that are based on their experience with commercial computer games. In that case, it is not unlikely that they will become disappointed, which may be reflected in less motivation and learning. Vogel et al. (2006) investigated the level of realism (photorealistic, high-quality cartoons, low-quality pictures, or unrealistic) in their meta-analysis but found no differences between the levels. The rapid technological developments and the increase in empirical studies in the last years justify a new examination of the impact of the level of realism.

Narrative. In game genres such as adventure games and role-playing games, narratives play an important role (Prensky, 2001). Research on learning from text shows that narratives foster learning and engagement. For example, compared with expository text or newspaper items, stories yield better recall, generate more inferences, and are more entertaining (Graesser, Singer, & Trabasso, 1994). Another argument for adding a narrative to the game is that it may scaffold problem solving during the game (Dickey, 2006). From a cognitive perspective, however, it can be argued that an engaging narrative may distract learners from the learning material and, given the limited cognitive capacity, withhold from them cognitive activities that yield learning (Mayer, Griffith, Naftaly, & Rothman, 2008). It is as yet unclear whether a narrative in a serious game will foster learning and engagement. Some value-added studies have been conducted, but they reported contradictory results. For example, McQuiggan, Rowe, Lee, and Lester (2008) found a negative effect of a narrative compared to a minimal narrative in a computer game, and Cordova and Lepper (1996) reported a beneficial effect when a narrative component (fantasy) was included.

Finally, we considered some methodological moderators. Meta-analyses allow a comparison of studies that use different experimental designs and different statistical methods. However, comparing studies with different degrees of methodological rigor may also obscure the results of the meta-analysis and thus jeopardize the conclusion that the weighted mean effect size is attributable to specific features of the studies and not to spurious factors that come with such studies. We identified three methodological indicators that potentially may influence the weighted mean effect size and the impact of study features.

Publication source. A potential danger in a meta-analysis is the "file drawer problem": the concern that the studies in the meta-analysis are not a correct reflection of all studies that are actually conducted (Ellis, 2010) because studies published in peer-reviewed journals and/or proceedings are more likely to have achieved statistical significance and larger effect sizes than are studies that have not been published (Rosenthal, 1995). Therefore, we used the publication source (peer-reviewed journal, proceedings, and unpublished) as a moderator variable.

Randomization. Second, we took into account whether a pure or a quasi-experimental design was used. In the latter case, participants are not randomized between the conditions, which may allow alternative explanations for the results that are found.

Experimental design. Finally, we considered whether a posttest-only design or a pretest–posttest design was used.

Moderator Variables for Motivation

For motivation, the same moderator variables were used, but we did not formulate hypotheses. We used the moderators to explore whether and to what extent contextual and situational factors have an impact on the motivational appeal of serious games.

Method

Literature Search

We started with computer-based searches via Google Scholar. The search terms we used were *game-based learning, PC games, video game, computer video game, serious games, educational games, simulation games, virtual environments,* and *muve.* If necessary, these search terms were combined with *learning, instruction, training, motivation,* and *engagement.* In addition, we investigated the references of previous meta-analyses and reviews on the effectiveness of serious games (Fletcher & Tobias, 2006; Ke, 2009; O'Neil et al., 2005; Sitzmann, 2011; Vogel et al., 2006; Wouters et al., 2009). In order to find unpublished but relevant studies, we asked researchers and educators within our network of scholars whether they were aware of relevant studies for the meta-analysis. Our meta-analysis covered the period from 1990 to 2012. Our research located 190 studies, of which 38 studies met our inclusion criteria (see the next section).

Inclusion Criteria and Coding

There were four inclusion criteria. First, the experimental group learned the content of the domain through a serious game, either as the sole instruction method or in combination with other instructional methods. In addition, there was a comparison group that engaged in an alternative instructional method. Second, the serious games and comparison groups had to receive the same learning content. Third, the study reported data or indications that allowed us to calculate or estimate effect sizes (group means and standard deviations, t test, F test, etc.). Fourth, we focused on nondisabled participants. The characteristics of each study that, in addition to the effect sizes and the sample sizes, were coded are described next.

Learning and retention. Two categories of learning outcomes were used to classify learning. "Knowledge" was used when a test involved knowledge of concepts, principles, definitions, symbols, or facts (e.g., Papastergiou, 2009, on computer knowledge). Studies in which learners had to solve problems, make decisions, or apply rules to a situation were coded as "Cognitive skills" (e.g., Kebritchi, Hirumi, & Bai, 2010). Retention was coded when a delayed measure for learning was available (the low number of pairwise comparisons does not allow a further breakdown in knowledge and cognitive skills). In the majority of the studies the delayed test took place 1 to 5 weeks after the intervention, but in one study the delayed test took place after 27 weeks (Segers & Verhoeven, 2003).

Motivation. We adopted a broad view on motivation. In the majority of the studies, a questionnaire or survey was used to measure motivation (e.g., Parchman, Ellis, Christinaz, & Vogel, 2000), interest (e.g., Ritterfeld, Shen, Wang, Nocera, & Wong, 2009), engagement (e.g., van Dijk, 2010), or attitude toward the topic involved in the experiment (e.g., Miller & Robertson, 2010). In one study, ratings of observed engagement (Brom, Preuss, & Klement, 2011) were used as a measure for motivation.

Learning arrangement of the comparison group. "Active instruction" refers to instruction methods that explicitly prompt learners to learning activities (e.g., exercises, hypertext training). We also coded whether the focus of the activity was drill-and-practice oriented or problem-solving oriented. "Passive instruction" includes listening to lectures; receiving classical instruction; and reading textbooks, expository text, or a PowerPoint presentation. Studies in which a combination of active and passive instruction was used were coded as "Mixed instruction." For example, Squire, Barnett, Grant, and Higginbotham (2004) used a comparison group with guided discovery involving interactive lectures, experiments, and demonstrations.

Serious game combined with other instructional methods. A study was coded "Inclusive" when the serious game was combined with other instructional methods (e.g., Kebritchi et al., 2010). When the serious game was the only instructional method, it was coded as "Exclusive" (e.g., Adams, Mayer, MacNamara, Koenig, & Wainess, 2012).

Number of training sessions. Studies in which learners engaged in only one training session with the serious game were coded "1 session." The time of this session ranged from 18 min (van Dijk, 2010) to 3 hr (Jong, Shang, Lee, Lee, & Law, 2006). Studies involving more than one session were coded as "Multiple sessions." The number of sessions varied from three (Gremmen & Potters, 1997) to 40 (Miller & Robertson, 2010).

Group size. When learners worked alone with the serious game, the study was coded "Individual." In the case of dyads or a larger group, the study was coded "Group." One study used a mix of individual and group game play (Kebritchi et al., 2010), but it was coded as Group.

Instructional domain. The studies in our research covered a broad range of domains. The domains biology, mathematics, language, and engineering were coded as such. Domains that were mentioned only a few times (aviation, computer science, geography, physics, military, and triage) were coded as "Other."

Age. The following categories were used: "Children" (until 12 years), "Preparatory education" (13 to 17 years), "Students" (18–24 years), and "Adults" (older than 25 years).

Level of realism. Games that were either textual (e.g., Gremmen & Potters, 1997) or schematic (e.g., mazelike type of games; see Papastergiou, 2009) were coded "Schematic." Cartoonlike games were coded as "Cartoon" (e.g., Brom et al., 2011), likewise photorealistic games were coded as "Realistic" (e.g., Kebritchi et al., 2010). In some studies, different types of serious games were used (e.g., Segers & Verhoeven, 2003). When the level of realism could be determined for all types, the study was coded under one of the aforementioned classifications (if all game types had the same level of realism) or as "Mixed" (if the game types had different levels of realism). "Unknown" was used when the level of realism of a game could not be determined.

Narrative. Games with a basic storyline (e.g., Ke, 2008) or more elaborated storyline (e.g., Barab et al., 2009) were coded as "Narrative." Games without a storyline were coded "Nonnarrative" (e.g., Cameron & Dwyer, 2005).

Methodological variables. For each study, three methodological variables were coded. To start with, the publication source could be a peer-reviewed journal, proceedings, or unpublished. Second, we coded whether or not participants were assigned randomly to the conditions. If schools or classes were randomly assigned to conditions and the learning effects were reported on an individual level, we coded the study as not random (e.g., Miller & Robertson, 2010). Finally, the experimental design of the study—either posttest only or pretest–posttest—was coded.

A random selection of 20 studies was coded independently by two raters. The mean intercoder agreement was 90.8%. Differences in codings were discussed until agreement was reached. The remaining studies were coded by the first author.

Calculation and Analysis of the Effect Sizes

Cohen's *d* was used as indicator for the effect size: The difference on the dependent variables (learning, retention, or motivation) between the serious game group and the comparison group was calculated and divided by the pooled standard deviation. When the means and/or standard deviations were not available, formulas were used to estimate the effect size based on data of the *t* test or the univariate *F* test (Glass, McGaw, & Smith, 1981), adjusted means, or gain scores (Hedges & Olkin, 1985).

Effect sizes for studies with small sample bias were corrected (cf. Hedges & Olkin, 1985). When multiple measurements were used for learning, retention, or motivation, an average was calculated. It was subsequently used to estimate the effect size. When multiple learning outcomes and/or multiple treatment or comparisons groups were used, each pairwise combination of a learning outcome and/or treatment or comparison group was treated as an independent study. The sample size was adjusted to avoid the overrepresentation of studies with multiple pairwise comparisons. For this purpose, we developed a procedure to assure that no comparison received an inappropriate weight (see the Appendix for a description of the procedure and an example).

We used the random-effects model for the main analyses and the moderator analyses with 95% confidence intervals around the weighted mean effect sizes. To calculate the effect sizes, we created a program in Excel using the formulas provided by Ellis (2010) and Borenstein, Hedges, Higgins, and Rothstein (2009).

Results

In total, 39 studies were identified; they yielded 77 pairwise comparisons on learning outcomes, 17 pairwise comparisons on retention, and 31 pairwise comparisons on motivation. Although we focused on studies conducted after 1990, 54% of the studies were conducted in the last 5 years (2007–2012). In total, 5,547 participants were involved. The sample sizes of the studies ranged from 16 to 1,105 participants. Table 1 (learning and retention) and Table 2 (motivation) present all included pairwise comparisons with effect sizes and their classification on the nonmethodological moderator variables.

The heterogeneity of effect sizes was confirmed only for learning (Q_{total} = 323.79, df = 76, $p < .001$) and for motivation (Q_{total} = 71.05, df = 30, $p < .001$) but not for retention (Q_{total} = 8.68, df = 16, $p > .05$). Therefore, a moderator analysis is justified for learning and motivation. For all analyses, alpha was set at .05.

Main Effect Analysis

The weighted mean effect sizes are presented in Table 3. Although we included a methodological moderator to examine a possible publication bias, we also calculated the fail-safe *N,* which is the number of studies averaging null results that has to be retrieved in order to reject the summary effect size. A publication bias is unlikely to occur when the fail-safe *N* (for this study, 3,489) exceeds the suggested threshold of the quintuple of pairwise comparisons plus 10 (Ellis, 2010), which is clearly the case in this review: $3{,}489 > 5 \times 77 + 10 = 395$.

The first hypothesis, which predicts that instruction with serious games yields higher learning gains than conventional instruction, is confirmed. The weighted mean effect size of 0.29 for learning in favor of serious games is statistically significant (z = 4.67, $p <$ .001). Also, the effect sizes of knowledge and cognitive skills show that serious games are superior to conventional instructional methods (knowledge: d = 0.27, z = 2.00, $p < .05$; cognitive skills: d = 0.29, z = 4.12, $p < .001$). The comparisons of the effect sizes of both learning outcomes reveal no significant differences ($p >$ .1). We also tested the homogeneity of effect sizes for the two learning outcomes. The Q_b statistic, $\chi^2(1) = 3.86$, $p > .1$, suggests that the differences between the two learning outcomes are attributable to sampling error. For this reason we used the overall learning effect size (d = 0.29) in the subsequent moderator analysis.

Table 1
Studies and Pairwise Comparisons of Serious Games vs. a Comparison Group and the Effects on Learning

Study	Adjusted *N*	Learning outcome	$d_{\text{immediate}}$	$d_{\text{retention}}$	Activity comparison group	Inclusive/exclusive	No. sessions	Group size	Domain	Age	Level of realism
Adams et al. (2012), Experiment 1	21	Knowledge	−1.34		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Realistic
	21	Skills	−0.36		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Realistic
Adams et al. (2012), Experiment 2											
Narrative	57	Skills	−0.31		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Realistic
Nonnarrative	57	Skills	−0.47		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Realistic
Annetta et al. (2009)	130	Skills	0.00		Active (practice)	Inclusive	1	Group	Biology	Preparatory education	Realistic
Bai et al. (2012)	219	Skills	0.19		Mixed	Inclusive	>1	Group	Math	Preparatory education	Realistic
Barab et al. (2009)											
Pilot study	35	Skills	0.66		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Realistic
Pilot study	35	Skills	0.87		Active (hypertext)	Exclusive	1	Individual	Biology	Students	Realistic
Main study	12	Skills	0.76		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Realistic
Main study	12	Skills	1.39		Passive (PowerPoint)	Exclusive	1	Group	Biology	Students	Realistic
Main study	12	Skills	0.29		Mixed	Exclusive	1	Individual	Biology	Students	Realistic
Main study	12	Skills	0.69		Mixed	Exclusive	1	Group	Biology	Students	Realistic
Barab et al. (2012)	65	Skills	0.47		Active (assignments)	Inclusive	>1	Individual	Biology	Preparatory education	Realistic
Barlett et al. (2009)	54	Skills	0.28		Active (Internet task)	Exclusive	1	Individual	Other	Students	Realistic
	59	Skills	0.19		Active (Internet task)	Exclusive	1	Individual	Other	Students	Realistic
Betz (1996)	24	Skills	—[a]	0.64	Passive (reading)	Inclusive	>1	Individual	Other	Students	Cartoon
Brom et al. (2011)	100	Knowledge	0.03	0.25	Passive (lecture)	Inclusive	1	Individual	Biology	Preparatory education	Cartoon
Cameron & Dwyer (2005)											
Game	140	Skills	0.21		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Schematic
Game with elaborated feedback	144	Skills	0.50		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Schematic
Game with responsive feedback	139	Skills	0.79		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Students	Schematic
Dede et al. (2005)	125	Knowledge	0.23		Mixed	Inclusive	>1	Individual	Biology	Preparatory education	Realistic
Gremmen & Potters (1997)	38	Skills	0.59	0.67	Passive (lecture)	Exclusive	>1	Group	Other	Students	Schematic
Jarvis & de Freitas (2009)	91	Skills	0.67		Active (practice)	Inclusive	1	Individual	Other	Adults	Realistic
Jong et al. (2006)	158	Skills	−0.09		Active (hypertext)	Inclusive	1	Individual	Math	Preparatory education	Cartoon
Ke (2008)	96	Skills	0.24		Active (practice)	Exclusive	>1	Mix	Math	Preparatory education	Cartoon
Ke & Grabowski (2007)											
Competitive	59	Skills	0.32		Active (practice)	Exclusive	>1	Individual	Math	Preparatory education	Cartoon
Cooperative	61	Skills	0.39		Active (practice)	Exclusive	>1	Group	Math	Preparatory education	Cartoon
Kebritchi et al. (2010)	193	Skills	0.27		Mixed	Inclusive	>1	Group	Math	Preparatory education	Realistic
Laffey et al. (2003)	56	Skills	0.96		Mixed	Inclusive	>1	Individual	Math	Children	Cartoon

(*table continues*)

Table 1 (*continued*)

Study	Adjusted N	Learning outcome	$d_{immediate}$	$d_{retention}$	Activity comparison group	Inclusive/exclusive	No. sessions	Group size	Domain	Age	Level of realism
McQuiggan et al. (2008)											
Narrative	77	Skills	−0.98		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Preparatory education	Realistic
Nonnarrative	73	Skills	−0.61		Passive (PowerPoint)	Exclusive	1	Individual	Biology	Preparatory education	Realistic
Miller & Robertson (2010)	42	Skills	0.46		Active (brain gym)	Inclusive	>1	Individual	Math	Children	Unknown
Miller & Robertson (2011)	635	Skills	0.08		Mixed	Inclusive	>1	Individual	Math	Children	Unknown
Moreno et al. (2001)											
Experiment 3	19	Knowledge	0.85		Active (multimedia)	Exclusive	1	Individual	Biology	Students	Schematic
Experiment 3	19	Skills	0.60		Active (multimedia)	Exclusive	1	Individual	Biology	Students	Schematic
Nicol & Anderson (2000)	12	Skills	−0.06	0.04	Active (exercises)	Exclusive	>1	Group	Math	Adults	Unknown
Okolo (1992)	41	Skills	0.00		Active (practice)	Exclusive	>1	Individual	Math	Students	Unknown
Papastergiou (2009)	88	Knowledge	0.61		Mixed	Inclusive	1	Individual	Other	Preparatory education	Schematic
Parchman et al. (2000)											
CBDP	10	Knowledge[b]	−0.91		Active (practice)	Exclusive	1	Individual	Engineering	Students	Cartoon
CBDP	10	Skills	0.32		Active (practice)	Exclusive	1	Individual	Engineering	Students	Cartoon
ECBI	15	Knowledge	−0.57		Mixed	Exclusive	1	Individual	Engineering	Students	Cartoon
ECBI	15	Skills	−0.47		Mixed	Exclusive	1	Individual	Engineering	Students	Cartoon
CI	15	Knowledge	0.21		Mixed	Exclusive	1	Individual	Engineering	Students	Cartoon
CI	15	Skills	−0.31		Mixed	Exclusive	1	Individual	Engineering	Students	Cartoon
Ricci et al. (1996)	30	Knowledge	−0.27	0.32	Active (test)	Exclusive	1	Individual	Other	Students	Schematic
	30	Knowledge	1.28	0.47	Passive (reading)	Exclusive	1	Individual	Other	Students	Schematic
Ritterfeld et al. (2009)	19	Knowledge	0.32	0.46	Active (hypertext)	Exclusive	1	Individual	Biology	Students	Realistic
	19	Skills	0.25	0.33	Active (hypertext)	Exclusive	1	Individual	Biology	Students	Realistic
	19	Knowledge	0.62	0.68	Passive (reading)	Exclusive	1	Individual	Biology	Students	Realistic
	19	Skills	0.28	0.15	Passive (reading)	Exclusive	1	Individual	Biology	Students	Realistic
Rosas et al. (2003)											
Reading	368	Skills	0.02		Mixed	Inclusive	>1	Individual	Language	Preparatory education	Cartoon
Spelling	368	Skills	−0.10		Mixed	Inclusive	>1	Individual	Language	Preparatory education	Cartoon
Mathematics	368	Skills	0.08		Mixed	Inclusive	>1	Individual	Math	Preparatory education	Cartoon
Seelhammer (2009)	20	Knowledge	−1.60		Passive (reading)	Inclusive	1	Individual	Biology	Preparatory education	Realistic
	20	Skills	−0.57		Passive (reading)	Inclusive	1	Individual	Biology	Preparatory education	Realistic
Segers & Verhoeven (2003)											
Immigrants, year 1	33	Knowledge	0.10	0.01	Mixed	Inclusive	>1	Individual	Language	Children	Cartoon
Immigrants, year 2	62	Knowledge	0.34	0.36	Mixed	Inclusive	>1	Individual	Language	Children	Cartoon
Natives, year 1	34	Knowledge	0.23	0.10	Mixed	Inclusive	>1	Individual	Language	Children	Cartoon
Natives, year 2	18	Knowledge	0.33	0.26	Mixed	Inclusive	>1	Individual	Language	Children	Cartoon
Sindre et al. (2009)	32	Knowledge	0.33		Active (exercises)	Inclusive	1	Individual	Other	Students	Unknown
	32	Knowledge	0.38		Passive (reading)	Inclusive	1	Individual	Other	Students	Unknown
Squire et al. (2004)	90	Skills	1.75		Mixed	Inclusive	>1	Individual	Other	Preparatory education	Cartoon
Suh et al. (2010)											
Listening	55	Skills	2.22		Mixed	Inclusive	>1	Group	Language	Preparatory education	Unknown
Speaking	55	Skills	0.22		Mixed	Inclusive	>1	Group	Language	Preparatory education	Unknown
Reading	55	Skills	1.20		Mixed	Inclusive	>1	Group	Language	Preparatory education	Unknown
Writing	55	Skills	1.71		Mixed	Inclusive	>1	Group	Language	Preparatory education	Unknown
van Dijk (2010)	24	Knowledge	−0.78	1.07	Passive (PowerPoint)	Exclusive	1	Individual	Other	Students	Realistic
	24	Knowledge	−0.33	1.06	Passive (PowerPoint)	Exclusive	1	Individual	Other	Students	Realistic

(*table continues*)

Table 1 (*continued*)

Study	Adjusted N	Learning outcome	$d_{immediate}$	$d_{retention}$	Activity comparison group	Inclusive/exclusive	No. sessions	Group size	Domain	Age	Level of realism
Van Eck & Dempsey (2002)											
Adventure/computer	23	Skills	0.08		Active (word problem)	Exclusive	1	Individual	Math	Preparatory education	Schematic
No adventure/computer	24	Skills	0.45		Active (word problem)	Exclusive	1	Individual	Math	Preparatory education	Schematic
Adventure/no computer	23	Skills	0.52		Active (word problem)	Exclusive	1	Individual	Math	Preparatory education	Schematic
No adventure/no computer	18	Skills	−0.28		Active (word problem)	Exclusive	1	Individual	Math	Preparatory education	Schematic
Virvou et al. (2005)											
Part 1	90	Knowledge	0.96		Mixed	Exclusive	>1	Individual	Other	Preparatory education	Realistic
Part 2, poor performance	30	Knowledge	1.79		Mixed	Exclusive	>1	Individual	Other	Preparatory education	Realistic
Part 2, medium performance	30	Knowledge	0.84		Mixed	Exclusive	>1	Individual	Other	Preparatory education	Realistic
Part 2, good performance	30	Knowledge	0.10		Mixed	Exclusive	>1	Individual	Other	Preparatory education	Realistic
Wrzesien (2010)	48	Knowledge	0.28		Passive (reading)	Exclusive	1	Group	Biology	Children	Realistic
Yip & Kwan (2006)	100	Knowledge	1.20		Active	Inclusive	>1	Individual	Language	Students	Unknown

Note. CBDP = computer-based drill and practice; ECBI = experimental computer-based instruction; CI = classical instruction.
[a] In the Betz (1996) study, learning was measured after a week. It was therefore classified as retention. [b] In the Parchman et al. (2000) study, knowledge of definitions and symbols was defined as knowledge and principle and rule application was defined as cognitive skills. The column Activity comparison group describes the instruction activity and how it was classified in the meta-analysis.

The second hypothesis predicts that instruction with serious games yields a higher level of retention than training with conventional instructional methods. Indeed, the results show that the superiority of serious games over conventional instructional methods is maintained in a delayed test (d = 0.36, z = 2.41, $p < .05$).

The third hypothesis predicts that serious games are more motivating than conventional instructional methods. Although the summary effect size is 0.26 in favor of serious games, the corresponding z score indicates that the difference in motivation is not statistically significant (z = 1.77, $p > .05$).

Moderator Analysis

We conducted a moderator analysis only for learning and motivation. For retention, a moderator analysis was not appropriate, given the homogeneous distribution of effect sizes (see the aforementioned Q_{total} statistic) and the low number of pairwise comparisons.

In order to compare the subgroups in the moderator analysis, we adopted the z-testing method for random effects with separate estimates of between-study variance (see Borenstein et al., 2009). When a moderator variable comprised more than two categories, the Holm–Bonferroni procedure was used to adjust the critical p value to control for the Type 1 error (cf. Ginns, 2005). In Holm's sequential version, the results of the Bonferroni tests are ordered from the smallest to the highest p value. The test with the lowest p value is then tested first with a Bonferroni correction involving all tests. The second test applies a Bonferroni correction involving one test less. This procedure continues for the remaining tests (Abdi, 2010).

Moderator analysis for learning. The results of the moderator analysis for learning are shown in the left part of Table 4.

The fourth hypothesis predicts that serious games yield more learning when the comparison group engages in passive instruction rather than active instruction. This hypothesis is not confirmed ($z_{active\text{-}passive} = -1.38$, $p > .05$). On the contrary, serious games do not improve learning more than does passive instruction. The beneficial effect of serious games is larger for mixed instructional methods than for passive instruction methods ($z_{mixed\text{-}passive} = 2.56$, $p = .005$). All other comparisons revealed no significant differences (ps $> .05$). Although the effect of serious games seems stronger for instruction with a focus on problem solving (d = 0.31) than for drill-and-practice-oriented instruction (d = 0.22), the difference is not significant ($z_{drill\text{-}and\text{-}practice\text{-}problem\ solving} = 0.45$, $p > .1$).

In Hypothesis 5, we expect that, for the experimental relative to the comparison group, serious games supplemented with other instructional methods will yield higher learning gains than serious games without supplemental instructional methods. The results confirm this hypothesis: Compared with conventional instruction methods, serious games yield higher learning gains irrespective of whether they are presented alone (d = 0.20) or supplemented with other instructional methods (d = 0.41), but learners learn most when serious games are supplemented with other instructional methods ($z_{inclusive\text{-}exclusive} = 1.66$, $p = .048$).

The sixth hypothesis predicts that multiple training sessions with serious games will yield higher learning gains than multiple training sessions with conventional instruction methods. When only one training session is involved, serious games are not more

Table 2
Studies and Pairwise Comparisons of Serious Games vs. a Comparison Group and the Effects on Motivation

Study	Adjusted *N*	Motivation	*d*	Activity comparison group	Inclusive/exclusive	No. sessions	Group size	Domain	Age	Level of realism
Annetta et al. (2009)	72	Observed engagement	0.81	Active (practice)	Inclusive	1	Group	Biology	Preparatory education	Realistic
Bai et al. (2012)	219	ARCS	0.30	Mixed	Inclusive	>1	Group	Math	Preparatory education	Realistic
Barab et al. (2012)	33	Question engagement	1.59	Active (assignments)	Inclusive	>1	Individual	Biology	Preparatory education	Realistic
Brom et al. (2012)	50	Question appeal	0.02	Passive (lecture)	Inclusive	1	Individual	Biology	Preparatory education	Cartoon
	50	Question value	−0.47	Passive (lecture)	Inclusive	1	Individual	Biology	Preparatory education	Cartoon
Ke (2008)	358	ATMI	0.47	Active (practice)	Exclusive	>1	Mix	Math	Preparatory education	Cartoon
Ke & Grabowski (2007)										
Competitive	59	ATMI	0.18	Active (practice)	Exclusive	>1	Individual	Math	Preparatory education	Cartoon
Cooperative	61	ATMI	0.55	Active (practice)	Exclusive	>1	Group	Math	Preparatory education	Cartoon
Kebritchi et al. (2010)	193	Survey	−0.23	Mixed	Inclusive	>1	Group	Math	Preparatory education	Realistic
Kuo (2007)	23	Survey	0.67	Active (explore)	Exclusive	>1	Individual	Biology	Children	Cartoon
	23	Visiting times	0.76	Active (explore)	Exclusive	>1	Individual	Biology	Children	Cartoon
Miller & Robertson (2010)	14	Learning self-concept	0.46	Active (brain gym)	Inclusive	>1	Individual	Math	Children	Unknown
	14	Self-esteem	0.13	Active (brain gym)	Inclusive	>1	Individual	Math	Children	Unknown
	14	Math self-concept	0.05	Active (brain gym)	Inclusive	>1	Individual	Math	Children	Unknown
Miller & Robertson (2011)	212	Learning self-concept	0.00	Mixed	Inclusive	>1	Individual	Math	Children	Unknown
	212	Self-esteem	−0.08	Mixed	Inclusive	>1	Individual	Math	Children	Unknown
	212	Math self-concept	0.12	Mixed	Inclusive	>1	Individual	Math	Children	Unknown
Moreno et al. (2001)										
Experiment 3	19	Question motivation	0.10	Active (multimedia)	Exclusive	1	Individual	Biology	Students	Schematic
Experiment 3	19	Question interest	0.24	Active (multimedia)	Exclusive	1	Individual	Biology	Students	Schematic
Papastergiou (2009)	88	Question	0.41	Mixed	Inclusive	1	Individual	Other	Preparatory education	Schematic
Parchman et al. (2000)	20	ARCS	0.01	Active (practice)	Exclusive	1	Individual	Engineering	Students	Cartoon
	30	ARCS	−0.72	Mixed	Exclusive	1	Individual	Engineering	Students	Cartoon
	31	ARCS	−0.12	Active (practice)	Exclusive	1	Individual	Engineering	Students	Cartoon
Ricci et al. (1996)	30	Question	0.57	Active (tests)	Exclusive	1	Individual	Other	Students	Schematic
	30	Question	1.21	Passive (reading)	Exclusive	1	Individual	Other	Students	Schematic
Ritterfeld et al. (2009)	38	Question	0.37	Active (hypertext)	Exclusive	1	Individual	Biology	Students	Realistic
	38	Question	0.34	Passive (text)	Exclusive	1	Individual	Biology	Students	Realistic
van Dijk (2010)	24	Question	0.25	Passive (PowerPoint)	Exclusive	1	Individual	Medicine	Students	Realistic
Wrzesien et al. (2010)	24	Question engagement	0.66	Passive (reading)	Exclusive	1	Group	Biology	Children	Realistic
	24	Question enjoyment	0.89	Passive (reading)	Exclusive	1	Group	Biology	Children	Realistic
	24	Question motivation	0.00	Passive (reading)	Exclusive	1	Group	Biology	Children	Realistic

Note. The column Activity comparison group describes the instruction activity and how it was classified in the meta-analysis (between brackets). ATMI = Attitude Towards Mathematics Inventory; ARCS = Attention Relevance Confidence Satisfaction.

Table 3
Main Effects for Learning, Retention, and Motivation Comparing Serious Games With Other Instructional Methods

Variable	*d*	*SE*	*k*	*N*	95% CI	Q_t
Learning	0.29**	.06	77	5,547	[0.17, 0.42]	323.97
Knowledge	0.27*	.14	25	948	[0.01, 0.54]	90.12
Skills	0.29**	.07	52	4,599	[0.15, 0.43]	226.61
Retention	0.36*	.16	16	499	[0.07, 0.68]	8.68
Motivation	0.26	.15	31	2,216	[−0.03, 0.56]	71.05

Note. d = weighted mean effect size (* $p > .05$; ** $p > .001$); *SE* = standard error of the effect size; k = number of pairwise comparisons; N = sum of the sample sizes of each pairwise comparison; CI = confidence interval; Q_t = homogeneity statistic.

effective than conventional instruction methods. However, consistent with the hypothesis, the results also show that multiple sessions yield higher learning gains for serious games than for conventional instruction methods ($d = 0.54$). Additionally, the comparison of the groups reveal that multiple sessions are more effective than only one session ($z_{1\ session\text{-}multiple\ sessions} = 3.94$, $p = .003$).

In Hypothesis 7, we predict that, for the experimental relative to the comparison group, learners learn more when they individually play serious games than when they play in a group. Not only do the results reject the hypothesis, but they also show that the reverse is the case: With serious games, both learners playing individually and those playing in a group learn more than the comparison group (respectively, $d = 0.22$ and $d = 0.66$), but learners who play serious games in a group learn more ($z_{individual\text{-}group} = 2.34$, $p = .01$).

In general, the results show that serious games improve learning more than conventional instruction methods in all domains except biology and engineering, but there is also much variation between the domains. Serious games are particularly effective in language ($d = 0.66$). For the experimental relative to the comparison group, serious games yield more learning in language than in biology ($z_{language\text{-}biology} = 2.28$, $p = .01$) and mathematics ($z_{language\text{-}math} = 2.25$, $p < .01$).

Serious games are superior to the comparison group for all age groups with the exception of adults. The comparisons of age groups reveal no differences (ps $> .1$). With respect to the level of realism, the results indicate that instruction with schematic serious games is superior to conventional instruction methods ($d = 0.46$). This is not true for cartoonlike or realistic serious games ($p > .05$). Mutual comparisons also show that schematic serious games are more effective than cartoonlike or realistic games ($z_{schematic\text{-}cartoonlike} = 1.89$, $p = .03$, $z_{schematic\text{-}realistic} = 2.25$, $p = .01$). Compared with conventional instruction methods, serious games without a narrative seem to be more effective than serious games with a narrative, but the difference is not significant ($z_{narrative\text{-}no\ narrative} = 1.34$, $p = .09$).

Turning to the methodological moderators, we see that only studies in peer-reviewed journals report higher learning gains for serious games. For proceedings and unpublished papers the effect sizes are even negative, but it should be noted that the number of pairwise comparisons in both publication sources is very low. Comparisons based on the Holm–Bonferroni procedure show no significant differences between the publication sources (ps $> .05$). The beneficial effect of serious games is contingent on the experimental rigor: Random assignment attenuates the effect of serious games ($z_{random\text{-}nonrandom} = 2.75$, $p < .003$). In fact, in studies with randomization, serious games are not more effective than conventional instruction methods. Finally, the experimental design of the study (posttest only: $d = 0.25$ vs. pretest–posttest design: $d = 0.32$) does not have an impact on the magnitude of the effect size ($z_{posttest\ only\text{-}preposttest} = 0.55$, $p > .1$).

Moderator analysis for motivation. The right side of Table 4 shows the moderator analysis for motivation. Two interesting observations can be made. First, serious games are more motivating compared to a group receiving active instruction ($d = 0.45$, $p = .02$). Second, relative to conventional instruction methods, serious games are more motivating when they are not combined with other instruction methods ($d = 0.37$, $p = .03$). We also found that, relative to a group receiving conventional instruction, schematic serious games are more motivating ($d = 0.51$, $p = .02$), but this conclusion is based on only five pairwise comparisons. All other moderators are not statistically significant. No comparisons of the subgroups within the moderator variables reached statistical significance (ps $> .1$).

Discussion

It is often argued that the affordances of computer games can be used to foster learning and motivation in instruction. Indeed several reviews have—at least partly—shown this potential (Ke, 2009; Vogel et al., 2006; Wouters et al., 2009), but the increase in empirical studies on serious games in the last 5 years justifies a new meta-analysis. In addition, it is still not clear which instructional and contextual factors have an impact on the effectiveness of serious games. Our results confirm the findings of the earlier reviews that, in general, serious games are more effective than conventional instruction methods. However, there are also some striking differences, which we discuss in this section.

Learning

The results on knowledge and cognitive skills suggest that training with serious games is more effective than training with conventional instruction methods. In line with Sitzmann (2011), the retention outcome shows that the cognitive gains are not attributable to the "freshness" of the learning material but that these gains persist in the long term. This retention effect is impor-

Table 4
Moderator Analysis Comparing Serious Games With Other Instructional Methods for Learning and Motivation

Variable	Learning *d*	Learning *k*	Learning 95% CI for *d*	Motivation *d*	Motivation *k*	Motivation 95% CI for *d*
Activity comparison group						
Active instruction	0.28***	24	[0.12, 0.45]	0.45*	13	[0.09, 0.80]
Drill and practice	0.22**	13	[0.08, 0.35]	0.27	9	[−0.20, 0.75]
Problem solving	0.31*	6	[0.01, 1.01]	0.88**	4	[0.31, 1.44]
Unknown	0.28*	5	[0.02, 0.54]			
Passive instruction	0.06	25	[−0.20, 0 .33]	0.24	12	[−0.32, 0.81]
Mixed	0.50***	28	[0.30, 0.70]	0.07	6	[−0.62, 0.76]
Computer game alone						
Inclusive	0.41***	29	[0.23, 0.59]	0.18	13	[−0.33, 0.69]
Exclusive	0.20*	48	[0.03, 0.37]	0.37*	18	[0.07, 0.67]
No. training sessions						
One session	0.10	47	[−0.07, 0.26]	0.26	17	[−0.21, 0.73]
Multiple sessions	0.54***	30	[0.35, 0.72]	0.26	14	[−0.13, 0.65]
Group size[a]						
Individual	0.22**	63	[0.09–0.36]	0.23	24	[−0.14, 0.59]
Group	0.66**	13	[0.32, 1.00]	0.35	6	[−0.31, 1.01]
Domain						
Biology	0.11	28	[−0.11, 0.33]	0.44	13	[−0.03, 0.91]
Math/arithmetic	0.17**	16	[0.07, 0.28]	0.15	11	[−0.30, 0.60]
Language	0.66**	11	[0.25, 1.07]	—	—	—
Engineering	−0.36	6	[−0.80, 0.09]	0.24	7	[−0.62, 1.10]
Others	0.54***	16	[0.23, 0.85]	—	—	—
Age						
Children	0.30**	8	[0.08, 0.52]	0.15	11	[−0.30, 0.60]
Preparatory education	0.33**	31	[0.13, 0.54]	0.32	10	[−0.15, 0.79]
Students	0.23*	36	[0.04, 0.42]	0.22	10	[−0.49, 0.93]
Adults	0.50	2	[−0.10, 1.10]	—	—	—
Level of realism						
Schematic	0.46***	14	[0.27, 0.65]	0.51*	5	[0.05, 0.96]
Cartoonlike	0.20	20	[−0.01, 0.40]	0.12	15	[−0.42, 0.67]
Photorealistic	0.14	32	[−0.08, 0.35]	0.40	9	[−0.16, 0.97]
Mixed	—	—	—	—	—	—
Unknown	0.72**	11	[0.27, 1.16]	0.71	2	[0.01, 1.42]
Narrative						
Yes	0.25***	62	[0.11, 0.39]	0.32	17	[−0.08, 0.71]
No	0.46**	15	[0.18, 0.73]	0.19	14	[−0.27, 0.64]
Methodological variables						
Publication source						
Peer-reviewed journal	0.36***	67	[0.24, 0.48]	0.24	28	[−0.07, 0.56]
Proceedings	−0.16	7	[−0.91, 0.58]	—	—	—
Unpublished	−0.20	3	[−0.83, 0.43]	0.55	3	[−0.10, 1.20]
Randomization						
Yes	0.08	35	[−0.13, 0.29]	0.36	8	[−0.15, 0.86]
No	0.44***	42	[0.29, 0.60]	0.25	23	[−0.10, 0.60]
Design						
Posttest only	0.25**	27	[0.07, 0.44]	0.12	9	[−1.02, 1.25]
Pre-posttest	0.32***	50	[0.16, 0.48]	0.30	22	[0.01, 0.60]

Note. d = weighted mean effect size ($^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$); k = number of pairwise comparisons; CI = confidence interval.
[a] Ke (2008) provided combined data only for cooperative and individualistic conditions. Therefore, this study is not considered in the user group variable.

tant, because it supports what teachers and instructors deem important: that serious games lead to well-structured prior knowledge on which learners can build on during their learning career.

The meta-analysis also distinguishes some situational and contextual factors. The positive effect of multiple training sessions on learning is larger for serious games than for conventional instruction methods. We assumed that the advantages of serious games would emerge when the players engaged in more training sessions and became used to the complex learning environment. However, the results also allow other explanations. For example, with respect to text comprehension, Kintsch, Welsch, Schmalhofer, and Zimny (1990) have shown that memory for the surface level and textbase-level representation of text decays over time, whereas memory for the situation model is robust to such decay. Perhaps immediately after learning from conventional instruction or the game, the textbase representation is still sufficiently available, causing no difference between the conventional instruction and game conditions. In contrast, after a decay of 2 to 4 days, students may need the situation model of the text to perform adequately on the test; then, the benefit by deeper processing in the game condition pays off (cf. Kintsch, 1998, p. 328). Some evidence for this assertion comes from the retention measure. Studies with a one-session learning stage in which an immediate and a delayed test is administered show no efficacy on the short term ($k = 9$, $d = .14$, $p < .1$), but they do in the long term ($k = 9$, $d = .40$, $p < .01$). However, some caution is warranted. It is possible that in the short term such brief training session cause worse learning and less motivation than other instruction methods, whereas in the long term positive effects may appear. For example, players may voluntarily play the game without being asked and in this way learn. It is also possible that they actually have learned but that the type of test that is administered (e.g., a knowledge test asking definitions of concepts) does not detect the deep level of knowledge. In this respect, we propose to include other methods to measure learning (Day, Arthur, & Gettman, 2001; Wouters, van der Spek, & van Oostendorp, 2011).

Our hypothesis predicting that serious games are more effective when the comparison group engages in passive instruction rather than in active instruction is not confirmed. On the contrary, serious games are not more effective than passive instruction. These results seem to contradict those of Sitzmann (2011), who found that simulation games were far more effective when compared with passive instruction than with active instruction. A closer examination of the results shows that this moderator confounds with the number of instruction sessions moderator, because almost all studies involving passive instruction are conducted during a learning stage of one session. In that case, the failure to find a positive effect of serious games over passive instruction may be attributable to the one-session learning stage. This conclusion is supported by a similar pattern for active instruction (1 session: $k = 16$, $d = 0.18$; >1 session: $k = 8$, $d = 0.43$) and mixed instruction (1 session: $k = 7$, $d = 0.15$; >1 session: $k = 21$, $d = 0.58$).

Another significant moderator is whether serious games are used as the only instructional method or are supplemented with other instructional methods. The meta-analysis shows that serious games are more effective when they are supplemented with other instructional methods than they are when used as sole instruction method. This may be due to the fact that game players in the latter case gain intuitive knowledge, but they are not prompted to verbalize the new knowledge and so do not anchor it more profound in their knowledge base (Leemkuil & de Jong, 2011; Wouters et al., 2008). The additional effect of supplemental instructional methods is that they prompt or support players to articulate the new knowledge and integrate it with their prior knowledge. These findings are also in line with other research showing that the active reflection or reviewing of information and experiences is beneficial for learning. For example, regarding pure versus guided discovery learning research has shown that learning by doing has to be supplemented with opportunities to reflect (cf. Mayer, 2004). Likewise, in the game cycle model of Garris et al. (2002), debriefing, defined as the review and analysis of events that occurred in the game itself, is regarded as the most critical part of the (serious) game experience. This finding is also useful from a practical point of view. Practitioners such as teachers are still reluctant to adopt serious games in the classroom. One of the perceptions is that it is difficult to integrate the serious game in their daily practice (cf. Baek, 2008), but the results show the potential of using serious games together with instruction methods that they already use in the classroom.

Contrary to our hypothesis, serious games are more effective when played in groups (in most studies the participants played in dyads) than when played alone. We proposed earlier that serious games foster some learning activities but that other learning activities, such as the articulation of knowledge, are not automatically addressed. These learning activities can be prompted by supplementing serious games with other instruction methods. The large effect of playing in a group suggests that this is also an effective method to incite these additional learning activities. However, this remains unclear, because most studies did not accurately describe the type of guidelines the players received for the collaboration. More research is needed, as well as a better understanding about the most effective group size (dyads, as in Annetta, Minogue, Holmes, & Chen, 2009, or many players, as in Suh, Kim, & Kim, 2010).

The results of the domain variable are difficult to interpret, because the variable confounds with other moderator variables. Remarkable is the large effect size for language. Rich multimodal environments such as computer games have characteristics that appear to be beneficial for language acquisition. For instance, graphics and dynamical visualizations may facilitate better encoding of meanings and interpretations of words (cf. dual coding theory; Clark & Paivio, 1991) or they may help learners to practice language in an authentic and playful way (e.g., the use of a massive multiplayer online role-playing game in Suh et al., 2010).

The results on the level of realism of serious games corroborate those of Vogel et al. (2006). They show that, from the perspective of learning, there is no argument to opt for photorealistic visual designs, because more basic designs such as schematic/textual and cartoonlike designs are equally or more effective. In that respect the results suggest that designers of serious games should focus more on the learning content and domain and less on visual design issues. It would be interesting to further categorize studies that we call photorealistic in photorealistic, 3-D, and virtual reality and to investigate how these levels of realism moderate learning. It would be particularly interesting when these new levels of realism are related to specific domains and types of knowledge. For example, are 3-D and virtual reality game environments more effective for learning a medical triage (the classification of victims) than a plain 2-D photorealistic game environment? For most age groups with the exception of adults, learning with serious games was more effective than conventional instruction. Vogel et al. (2006) did not find a difference between children and adults. They speculated that this was somewhat counterintuitive, given the fact that children have shorter attention spans and lower intrinsic motivation and thus may learn better than adults with computer games. Although we observed a difference, it is premature to draw a conclusion because the adults age group comprised only two comparisons.

Although serious games with a narrative are not more effective than serious games without a narrative when compared with a conventional instruction method, the difference does suggest that including a narrative is counterproductive. In this respect it seems to support the argument of Adams et al. (2012) that players will use too much of their cognitive capacity for processing the narrative information that is not directly related to the learning content. We concur with them that a story with a theme that is closely related to the learning goals may improve the effect of a narrative. Assuming that a narrative consists of a series of related events (e.g., an initiating event, exposition, complication, climax, and resolution; see Brewer & Lichtenstein, 1982), the manipulation of the order of these events may also trigger relevant cognitive processes. The role of the manipulation of narrative events in games is still unexplored, but research on texts has shown that the introduction of surprise can be effective in terms of recall of story information and appreciation of the story (Hoeken & van Vliet, 2000). Some support for the effective use of surprising events in serious games comes from van der Spek (2011), who had learners play a narrative-based serious game in which they learned how to apply a medical procedure. During the game, specifically designed surprising events were triggered, and learners could not rely anymore on the procedure that they had learned. For example, due to a sudden failure in a power box, there was not sufficient light to perform a necessary step in the procedure. It was hypothesized that this would force players to rethink the medical procedure they had used before and to develop another solution in order to perform that step (see also Kintsch, 1980). Indeed, the unexpected events yielded a higher level of deep knowledge without a decline in the reported engagement.

We did not find statistically significant evidence for a publication bias. This is in contrast with the strong publication bias found by Sitzmann (2011) for simulation games. Possibly, the small number of unpublished pairwise comparisons (three comparisons from three studies) in our meta-analysis complicates the detection of a publication bias. We did find some evidence that the methodological rigor of the studies moderates the magnitude of the effect sizes: Designs with randomization of participants report significantly smaller effect sizes in favor of serious games than do studies with no randomization.

Motivation

Perhaps the foremost reason to use serious games is their alleged motivational appeal (Garris et al., 2002; Malone, 1981). The assumption underlying the motivational appeal of serious games is based on the high entertainment value of commercial computer games. However, the results of the meta-analysis show that serious games are not more motivating than the instructional methods used in the comparison group (d = 0.26, but the difference is not significant). Three plausible arguments may explain the lack of higher motivation for serious games. To start with, it is possible that serious games are not more motivating than other instructional methods. Reasoning from the self-determination approach, Ryan et al. (2006) have argued that autonomy supports intrinsic motivation. Consequently, conditions that limit the sense of control or freedom of action may undermine intrinsic motivation (Deci, Koestner, & Ryan, 1999). In serious games, the level of control is twofold: It is applicable to actions and decisions within the game but also to the instructional context, where decisions about issues such as the type of game and when to play the game have to be made. It is relevant to investigate whether variations in the level of control that serious games offer moderate intrinsic motivation. We tried to classify these variations in the studies included in this meta-analysis but found that the majority of the papers lacked sufficient information for us to do this adequately. With respect to the level control in the instructional context, an essential difference between leisure computer games and serious games is that the former are chosen by the players and played whenever and for as long as they want, whereas the type of game that is used and the playing time are generally defined by the curriculum in the case of serious games. Within the instructional context, it is possible that the lack of control on these decisions has attenuated the motivation appeal of serious games.

The second explanation contends that the connection between game design with a focus on entertainment and instructional design with a focus on learning is not a natural one. Several dimensions that have to be resolved in order to create really engaging serious games, such as learning versus playing or freedom versus control, have been outlined (de Castell & Jenson, 2003; Wouters, van Oostendorp, Boonekamp, & van der Spek, 2011). Take the situation in which a designer uses a pop-up screen with a message that prompts the player to reflect. From an instructional design perspective such a focus may yield learning, but it is also likely that such an intervention will disturb the flow of the game and consequently undermine the entertaining nature of the game. It is plausible that the lack of motivational appeal is a reflection of the fact that the world of game design and that of instructional design are not yet integrated. If this is true, more research on factors that connect the worlds of game design and instructional design is required. Interesting in this respect is the work of Habgood and Ainsworth (2011), who found that the integration of arithmetical content with the game mechanics that make playing games entertaining was more motivating than a game version in which both components were not integrated. The third explanation stems from an examination of the methods that are commonly used for the measurement of motivation. The question can be raised whether it makes sense to measure affective states such as motivation and enjoyment with questionnaires and surveys after game play; physiological or behavioral measures such as eye tracking and skin conductance seem to be more appropriate methods, because they can be collected during game play. Also, the player's motivation during game play may be attenuated after the game has finished. In 30 of the 31 pairwise comparisons in the meta-analysis, motivation was measured with a survey or questionnaire conducted after game play. The exception (Annetta et al., 2009) used the rating of observed engagement during game play as motivation measurement and found that the game was more motivating than the instructional treatment of the comparison group who received practice and group discussion (estimated effect size d = 0.81).

Limitations and Directions for Future Research

Scholars have different views on what studies to include in a meta-analysis, varying from a broad sample with different study characteristics coded to a restricted sample that meets specific criteria. In this meta-analysis, we have chosen a broad focus including not only studies conducted in controlled laboratory settings but also studies that took place in a classroom setting. At the same time we have tried to further qualify the weighted mean effect size of the analysis with a number of moderators such as the methodological quality of the studies or the distribution of learning (one session vs. multiple sessions). We are aware of the fact that another view on what studies to include in the meta-analysis may lead to other conclusions regarding the effectiveness of serious games. For example, if only studies with a randomized sample and a pretest–posttest design are considered, the positive effect in favor of serious games may disappear. In addition, our selection of moderators is not exhaustive, and other interesting features of studies (e.g., gender) may influence the effect size.

A broad range of serious games, from adventure games to puzzle games, and their application in different domains have been examined. This large variation justifies some caution when generalizing the results. The same domain can be approached from different game genres. For example, Kebritchi et al. (2010) used a sophisticated 3-D adventure game to teach mathematical skills, and Van Eck and Dempsey (2002), in the same domain, used a basic simulation game. Despite the different game genres that were used, both studies contributed to the $d = 0.17$ for mathematics. It would be interesting to investigate whether specific game genres (e.g., adventure games, simulation games) are more apt to teach specific domains (e.g., mathematics).

Our results corroborate other findings indicating that serious games are a more effective than other instruction methods (cf. Sitzmann, 2011; Vogel et al., 2006). The next step is more value-added research on specific game features that determine this effectiveness. Given the increasing number of empirical studies with serious games, we believe, a meta-analysis on serious game features can be successful. An example is the role of competition, which is regarded by some scholars as a crucial characteristic of computer games (see the introduction), but the question is whether competition is required to make effective and compelling serious games. Our review of the literature revealed some studies comparing competition and noncompetition game versions (Ke, 2008; Ke & Grabowski, 2007; Van Eck & Dempsey, 2002) that warrant such an investigation. Also, from a cognitive consequences approach, there are interesting directions for future research. For example, we found many studies investigating the effect of playing computer games on basic cognitive abilities, such as visual attention and spatial ability. We did not take these studies into account, because the "no activity" control group in these studies did not meet our inclusion criteria. With a sample of 17 comparisons we found a $d = 0.33$, indicating that computer games are effective to train basic cognitive skills. Assuming that these basic cognitive skills are associated with cognitive skills such as problem solving, it would be interesting to examine whether serious games foster these cognitive processes and whether training of these processes also yields a better performance on cognitive skills such as problem solving.

Besides the issue of the method of measurement of motivation that we addressed earlier, the definition of motivation should be examined. We applied a broad definition of motivation, which included engagement, interest, enjoyment, the ARCS (Attention Relevance Confidence Satisfaction; see Bai et al., 2012) and ATMI (Attitude Towards Mathematics Inventory; see Ke, 2008) scales, and the attitude of the player toward school or a school domain. The question can be raised whether all these definitions indeed refer to motivation or whether they represent different constructs. For example, to what extent does attitude toward school (Miller & Robertson, 2010) reflect dimensions of the construct motivation?

The conclusion that we have drawn from these results is that specific instructional or contextual features, such as supplementing with other instructional methods and working in groups, increase the effect of serious games. We have suggested that these features may have enabled learners to engage in learning activities from which they would otherwise refrain. More research is required if these features indeed foster additional learning activities (e.g., with think-aloud protocols). And, if this is true, can we design serious games in such a way that these learning activities are also activated in stand-alone serious games or when learners play solitary games? In other words, can we design serious games in such a way that players are automatically prompted to reflect on their performance during game play?

References

References marked with an asterisk indicate studies included in the meta-analysis.

Abdi, H. (2010). Holm's sequential Bonferroni procedure. In N. Salkind (Ed.), *Encyclopedia of research design* (pp. 574–578). Thousand Oaks, CA: Sage.

*Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104,* 235–249. doi:10.1037/a0025595

*Annetta, L. A., Minogue, J., Holmes, S. Y., & Chen, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53,* 74–85. doi:10.1016/j.compedu.2008.12.020

Baek, Y. K. (2008). What hinders teachers in using computer and video games in the classroom? Exploring factors inhibiting the uptake of computer and video games. *CyberPsychology & Behavior, 11,* 665–671. doi:10.1089/cpb.2008.0127

*Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology, 43,* 993–1003. doi:10.1111/j.1467-8535.2011.01269.x

Bandura, A. (1976). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

*Barab, S., Pettyjohn, P., Gresalfi, M., Volk, C., & Solomou, M. (2012). Game-based curriculum and transformational play: Designing to meaningfully positioning person, content, and context. *Computers & Education, 58,* 518–533. doi:10.1016/j.compedu.2011.08.001

*Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., & Warren, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology, 18,* 305–320. doi:10.1007/s10956-009-9171-5

*Barlett, C. P., Vowels, C. L., Shanteau, J., Crow, J., & Miller, T. (2009). The effect of violent and non-violent computer games on cognitive performance. *Computers in Human Behavior, 25,* 96–102. doi:10.1016/j.chb.2008.07.008

*Betz, J. A. (1996). Computer games: Increase learning in an interactive multidisciplinary environment. *Journal of Educational Technology Systems, 24,* 195–205. doi:10.2190/119M-BRMU-J8HC-XM6F

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.

Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics, 6,* 473–486. doi:10.1016/0378-2166(82)90021-2

*Brom, C. M., Preuss, M., & Klement, D. (2011). Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. *Computers & Education, 57,* 1971–1988. doi:10.1016/j.compedu.2011.04.007

*Cameron, B., & Dwyer, F. (2005). The effect of online gaming, cognition and feedback type in facilitating delayed achievement of different learning objectives. *Journal of Interactive Learning Research, 16,* 243–258.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3,* 149–210. doi:10.1007/BF01320076

Clark, R. E., Yates, K., Early, S., & Moulton, K. (2010). An analysis of the failure of electronic media and discovery-based learning: Evidence for the performance benefits of guided training methods. In K. H. Silber & R. Foshay (Eds.), *Handbook of training and improving workplace performance: Vol. I. Instructional design and training delivery* (pp. 263–297). Somerset, NJ: Wiley.

Cordova, D. L., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88,* 715–730. doi:10.1037/0022-0663.88.4.715

Day, E. A., Arthur, W., Jr., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86,* 1022–1033. doi:10.1037/0021-9010.86.5.1022

de Castell, S., & Jenson, J. (2003). Serious play. *Journal of Curriculum Studies, 35,* 649–665. doi:10.1080/0022027032000145552

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125,* 627–668. doi:10.1037/0033-2909.125.6.627

*Dede, C., Clarke, J., Ketelhut, D. J., Nelson, B., & Bowman, C. (2005, April). *Students' motivation and learning of science in a multi-user virtual environment*. Paper presented at the meeting of the American Educational Research Association, Montréal, Quebec, Canada.

Dickey, M. D. (2006). Game design narrative for learning: Appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. *Educational Technology Research and Development, 54,* 245–263. doi:10.1007/s11423-006-8806-y

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, England. Cambridge University Press.

Fletcher, J. D., & Tobias, S. (2006, February). *Using computer games and simulations for instruction: A research review*. Paper presented at the Conference on New Learning Technologies, Society for Applied Learning Technology, Orlando, FL.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33,* 441–467. doi:10.1177/1046878102238607

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15,* 313–331. doi:10.1016/j.learninstruc.2005.07.001

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Thousand Oaks, CA: Sage.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395. doi:10.1037/0033-295X.101.3.371

*Gremmen, H., & Potters, J. (1997). Assessing the efficacy of gaming in economic education. *Journal of Economic Education, 28,* 291–303.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20,* 169–206. doi:10.1080/10508406.2010.508029

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hoeken, H., & van Vliet, M. (2000). Suspense, curiosity, and surprise: How discourse structure influences the affective and cognitive processing of a story. *Poetics, 27,* 277–286. doi:10.1016/S0304-422X(99)00021-2

Inkpen, K., Booth, K. S., Klawe, M., & Upitis, R. (1995). *Playing together beats playing apart, especially for girls*. Paper presented at the meeting of Computer Support for Collaborative Learning '95, Bloomington, IN.

*Jarvis, S., & de Freitas, S. (2009). Evaluation of an immersive learning programme to support triage training. *Proceedings of the First IEEE International Conference in Games and Virtual Worlds for Serious Applications* (pp. 117–122). doi:10.1109/VS-GAMES

*Jong, M. S. Y., Shang, J. J., Lee, F. L., Lee, J. H. M., & Law, H. Y. (2006). Learning online: A comparative study of a situated game-based learning approach and a traditional web-based approach. In Z. Pan, R. Aylett, H. Diener, X. Jin, S. Göbel, & L. Li (Eds.), *Lecture notes in computer science: Vol. 3942. Technologies for e-learning and digital entertainment* (pp. 541–551). Berlin, Germany: Springer.

*Ke, F. (2008). Computer games application within alternative classroom goal structures: Cognitive, metacognitive, and affective evaluation. *Education Technology Research and Development, 56,* 539–556. doi:10.1007/s11423-008-9086-5

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (Vol. 1, pp. 1–32). Hershey, PA: Information Science Reference.

*Ke, F., & Grabowski, B. (2007). Gameplaying for maths learning: Cooperative or not? *British Journal of Educational Technology, 38,* 249–259. doi:10.1111/j.1467-8535.2006.00593.x

*Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effect of modern computer games on mathematics achievement and class motivation. *Computers & Education, 55,* 427–443. doi:10.1016/j.compedu.2010.02.007

Kintsch, W. (1980). Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics, 9,* 87–98. doi:10.1016/0304-422X(80)90013-3

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159. doi:10.1016/0749-596X(90)90069-C

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78,* 311–328. doi:10.1037/0021-9010.78.2.311

*Kuo, M.-J. (2007, March). How does an online game based learning environment promote students' intrinsic motivation for learning natural science and how does it affect their learning outcomes? In T.-W. Chan, A. Paiva, D. W. Shaffer, Kinshuk, & J.-C. Yang (Eds.), *The First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning, 2007. DIGITEL '07,* 135–142. doi:10.1109/DIGITEL.2007.28

*Laffey, J. M., Espinosa, L., Moore, J., & Lodree, A. (2003). Supporting learning and behavior of at-risk young children: Computers in urban education. *Journal of Research on Technology in Education, 35,* 423–440.

Leemkuil, H., & de Jong, T. (2011). Instructional support in games. In S. Tobias & D. Fletcher (Eds.), *Computer games and instruction* (pp. 353–369). Charlotte, NC: Information Age.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5,* 333–369. doi:10.1207/s15516709cog0504_2

Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59,* 14–19. doi:10.1037/0003-066X.59.1.14

Mayer, R. E. (2011). Multimedia learning and games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 281–305). Charlotte, NC: Information Age.

Mayer, R. E., Griffith, E., Naftaly, I., & Rothman, D. (2008). Increased interestingness of extraneous details leads to decreased learning. *Journal of Experimental Psychology: Applied, 14,* 329–339. doi:10.1037/a0013835

*McQuiggan, S., Rowe, J., Lee, S., & Lester, J. (2008). *Story-based learning: The impact of narrative on learning experiences and outcomes*. Montreal, Quebec, Canada.

*Miller, D. J., & Robertson, D. P. (2010). Using a games console in the primary classroom: Effects of brain training programme on computation and self-esteem. *British Journal of Educational Technology, 41,* 242–255. doi:10.1111/j.1467-8535.2008.00918.x

*Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology, 42,* 850–864. doi:10.1111/j.1467-8535.2010.01114.x

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology, 97,* 117–128. doi:10.1037/0022-0663.97.1.117

*Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents. *Cognition and Instruction, 19,* 177–213. doi:10.1207/S1532690XCI1902_02

*Nicol, M. M., & Anderson, A. (2000). Computer-assisted vs. teacher-directed teaching of numeracy in adults. *Journal of Computer Assisted Learning, 16,* 184–192. doi:10.1046/j.1365-2729.2000.00131.x

*Okolo, C. M. (1992). The effect of computer-assisted instruction format and initial attitude on the arithmetic facts proficiency and continuing motivation of students with learning disabilities. *Exceptionality, 3,* 195–211. doi:10.1080/09362839209524815

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *Curriculum Journal, 16,* 455–474. doi:10.1080/09585170500384529

*Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52,* 1–12. doi:10.1016/j.compedu.2008.06.004

*Parchman, S. W., Ellis, J. A., Christinaz, D., & Vogel, M. (2000). An evaluation of three computer-based instructional strategies in basic electricity and electronic. *Military Psychology, 12,* 73–87. doi:10.1207/S15327876MP1201_4

Pierfy, D. A. (1977). Comparative simulation game research: Stumbling blocks and stepping. *Simulation & Games, 8,* 255–268. doi:10.1177/003755007782006

Prensky, M. (2001). *Digital game-based learning*. New York, NY: McGraw-Hill.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14,* 154–166. doi:10.1037/a0019440

Renkl, A., & Atkinson, R. K. (2002). Learning from examples: Fostering self-explanations in computer-based learning environments. *Interactive Learning Environments, 10,* 105–119. doi:10.1076/ilee.10.2.105.7441

*Ricci, K. E., Salas, E., & Cannon-Bowers, J. A. (1996). Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology, 8,* 295–307. doi:10.1207/s15327876mp0804_3

*Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Wong, W. L. (2009). Multimodality and interactivity: Connecting properties of serious games with educational outcomes. *CyberPsychology & Behavior, 12,* 691–697. doi:10.1089/cpb.2009.0099

*Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., . . . Salinas, M. (2003). Beyond Nintendo: Design and assessment of educational video games for first and second grade students. *Computers & Education, 40,* 71–94. doi:10.1016/S0360-1315(02)00099-4

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118,* 183–192. doi:10.1037/0033-2909.118.2.183

Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 271–286). New York, NY: Cambridge University Press.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30,* 347–365. doi:10.1007/s11031-006-9051-8

*Seelhammer, C., & Niegemann, H. M. (2009). Playing games to learn—Does it actually work? *Proceedings of the 17th International Conference on Computers in Education* (pp. 675–681). Hong Kong: Asia-Pacific Society for Computers in Education.

*Segers, E., & Verhoeven, L. (2003). Effects of vocabulary training by computer in kindergarten. *Journal of Computer Assisted Learning, 19,* 557–566. doi:10.1046/j.0266-4909.2003.00058.x

*Sindre, G., Natvig, L., & Jahre, M. (2009). Experimental validation of the learning effect for a pedagogical game on computer fundamentals. *IEEE Transactions on Education, 52,* 10–18. doi:10.1109/TE.2007.914944

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64,* 489–528. doi:10.1111/j.1744-6570.2011.01190.x

*Squire, K., Barnett, B., Grant, J. M., & Higginbotham, T. (2004). Electromagnetism supercharged! Learning physics with digital simulation games. *Proceedings of the International Conference on Learning Sciences, 6,* 513–520.

*Suh, S., Kim, S. W., & Kim, N. J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *Journal of Computer Assisted Learning, 26,* 370–378. doi:10.1111/j.1365-2729.2010.00353.x

Tobias, S., Fletcher, J. D., Dai, D. Y., & Wind, A. P. (2011). Review of research on computer games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 127–222). Charlotte, NC: Information Age.

van der Meij, H., Albers, E., & Leemkuil, H. (2011). Learning from games: Does collaboration help? *British Journal of Educational Technology, 42,* 655–664. doi:10.1111/j.1467-8535.2010.01067.x

van der Spek, E. D. (2011). *Experiments in serious game design: A cognitive approach* (Unpublished doctoral dissertation). Universiteit Utrecht, Utrecht, the Netherlands.

*van Dijk, V. (2010). *Learning the triage procedure: Serious gaming based on guided discovery learning versus studying worked examples* (Unpublished master's thesis). Universiteit Utrecht, Utrecht, the Netherlands.

*Van Eck, R., & Dempsey, J. (2002). The effect of competition and contextualized advisement on the transfer of mathematics skills in a computer-based instructional simulation game. *Educational Technology, Research and Development, 50,* 23–41. doi:10.1007/BF02505023

*Virvou, M., Katsionis, G., & Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness. *Educational Technology & Society, 8*(2), 54–65.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34,* 229–243. doi:10.2190/FLHV-K4WA-WPVQ-H0YM

Wittrock, M. C. (1974). Learning as a generative activity. *Educational Psychologist, 11,* 87–95. doi:10.1080/00461527409529129

Wouters, P., van Oostendorp, H., Boonekamp, R., & van der Spek, E. D. (2011). The role of game discourse analysis and curiosity in creating engaging and effective serious games by implementing a back story and foreshadowing. *Interacting With Computers, 23,* 329–336. doi:10.1016/j.intcom.2011.05.001

Wouters, P., Paas, F., & van Merriënboer, J. J. M. (2008). How to optimize learning from animated models: A review of guidelines base on cognitive load. *Review of Educational Research, 78,* 645–675. doi:10.3102/0034654308320320

Wouters, P., van der Spek, E. D., & van Oostendorp, H. (2009). Current practices in serious game research: A review from a learning outcomes perspective. In T. M. Connolly, M. Stansfield, & L. Boyle (Eds.), *Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices* (pp. 232–250). Hershey, PA: IGI Global.

Wouters, P., van der Spek, E. D., & van Oostendorp, H. (2011). Measuring learning in serious games: A case study with structural assessment. *Educational Technology Research and Development, 59,* 741–763. doi:10.1007/s11423-010-9183-0

*Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education, 55,* 178–187. doi:10.1016/j.compedu.2010.01.003

*Yip, F. W. M., & Kwan, A. C. (2006). Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International, 43,* 233–249. doi:10.1080/09523980600641445

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38*(9), 25–32. doi:10.1109/MC.2005.297

Appendix

Procedure to Adjust Sample Size

In the formula $N_{\text{adjusted}} = [(N_{\text{experimental group}}/a) + (N_{\text{comparison group}})/b]/c$, a is the number of comparison groups, b is the number of experimental groups with a serious game, and c is the number of dependent variables. For example, Parchman, Ellis, Christinaz, and Vogel (2000) used two different learning outcomes ($c = 2$), three comparison groups ($a = 3$), and one experimental group ($b = 1$). The number of participants was 20 in the experimental group, 13 in the drill-and-practice comparison group, 23 in the computer-based instruction comparison group, and 24 in the classical instruction comparison group. This means that a total of 80 learners participated in this study. The combination of learning outcomes and comparison groups yields six pairwise comparisons. For each pairwise comparison, an adjusted n was calculated based on the number of participants in the experimental and comparison groups. As shown in Table A1, the sum of the adjusted n of all pairwise comparisons equaled the total number of participants of that study.

Table A1
Example of Adjustment of Sample Size With Two Different Learning Outcomes and Three Comparison Groups

Pairwise comparison	*N* experimental group	*N* comparison group	Dependent variable	Formula	Adjusted *n*
1	20	13	Knowledge	([20/3] + [13/1])/2	9.83
2	20	13	Skills	([20/3] + [13/1])/2	9.83
3	20	23	Knowledge	([20/3] + [23/1])/2	14.38
4	20	23	Skills	([20/3] + [23/1])/2	14.38
5	20	24	Knowledge	([20/3] + [24/1])/2	15.33
6	20	24	Skills	([20/3] + [24/1])/2	15.33
				Total *N*	80

Received November 1, 2011
Revision received October 12, 2012
Accepted November 12, 2012 ■